CORE
Metadata, citation and similar papers at core.ac.uk
Provided by HAL-Univ-Nantes

Mode-dependent transform competition for HEVC

Adrià Arrufat, Pierrick Philippe, Olivier Déforges

To cite this version:

Adrià Arrufat, Pierrick Philippe, Olivier Déforges. Mode-dependent transform competition for HEVC. Image Processing (ICIP), 2015 IEEE International Conference on, Sep 2015, Québec, Canada. IEEE ICIP (Image Processing) 2015, pp.1598-1602, 2015, <10.1109/ICIP.2015.7351070>. <hal-01244783>

HAL Id: hal-01244783

https://hal.archives-ouvertes.fr/hal-01244783

Submitted on 16 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MODE-DEPENDENT TRANSFORM COMPETITION FOR HEVC

Adrià Arrufat, Pierrick Philippe[*]

Orange Labs
4, Rue du Clos Courtel
35512 Cesson-Sévigné — FRANCE

Olivier Déforges

IETR/INSA
UMR CNRS 6164
35043 Rennes — FRANCE

ABSTRACT

Transform coding plays a key role in state-of-the-art video coders, such as HEVC. However, transforms used in current solutions do not cover the varieties of video coding signals. This work presents an adaptive transform design method that enables the use of multiple transforms in HEVC. A different transform set is learnt for each intra prediction mode, allowing the video encoder to perform better decisions regarding block sizes, prediction modes and transforms. Different systems are proposed to accommodate trade-offs between complexity and performance. Bit rate reductions in the range of 2% to 7% are reported, depending on complexity.

Index Terms— HEVC, transform coding, KLT, MDDT, rate-distortion optimised transforms

1. INTRODUCTION

Block-based transform coding is used in current lossy compression schemes, such as High Efficiency Video Coding (HEVC) [1] to compact the signal energy. Despite being an important step in video coding, employed transforms are generic and not adapted to particular signal statistics: trigonometric transforms in the DCT family are still considered in state-of-the-art video standards such as HEVC.

Some research has been conducted in order to determine transforms better adapted to the different underlying varieties of video signal statistics. In this context the mode-dependent directional transform (MDDT) [2, 3, 4] has been studied over the recent years. The idea behind the MDDT is to design transforms adapted to each intra prediction (IP) mode, which essentially implies the design of a Karhune-Loève transform (KLT) per prediction mode.

In this paper, the MDDT concept is extended through the usage of a set of transforms adapted to each IP mode: for a given prediction mode, the encoder selects the best transform in the provided set. The design of multiple transforms and their implementation into HEVC have already been presented in our previous work [5, 6], following two different approaches:

- In [5], it is shown that a transform design method called rate-distortion optimised transform (RDOT) yields to better coding efficiency compared to the traditional KLT design. This method is used to design transforms adapted for each IP mode as in the MDDT system. The new generated transforms replace the default HEVC transforms on targeted block sizes.
- Work in [6] follows a different approach; instead of designing one transform for each IP mode and replace the default transforms, it provides complementary transforms to HEVC core transforms that are shared amongst all prediction modes. The HEVC encoder selects the best performing transform in a rate-distortion optimisation (RDO) loop, along with the most adapted prediction mode and block size.

Both approaches suggest the use of non-separable transforms and bit-rate reductions of around 2% are reported using either technique: this improvement comes at the cost of significant increase in complexity and storage needs.

In this paper, a combination of both approaches is proposed. It is shown that an increase in performance can be achieved, while reducing the algorithmic complexity.

This paper is organised as follows. Section 2 introduces the design of separable and non-separable RDOTs. The combined approach is presented in section 3, where multiple transforms are provided for each IP mode. This approach is implemented in HEVC and the performance results are presented and discussed in section 4.

2. RATE-DISTORTION OPTIMISED TRANSFORMS

Sparse orthogonal transforms, or rate-distortion optimised transforms (RDOTs), were introduced in [7] by Sezer et al. and are explained more in detail in [8]. The RDOTs provide better compression gains than the KLT in the context of video coding as reported in [5].

A RDOT is a transform A_{opt} that minimises the following quantity:

$$\mathbf{A}_{opt} = \arg\min_{\mathbf{A}} \sum_{\forall i} \min_{\mathbf{c}_i} \left(\|\mathbf{x}_i - \mathbf{A}^T \mathbf{c}_i\|_2^2 + \lambda \|\mathbf{c}_i\|_0 \right) \quad (1)$$

[*]This author performed the work while at B-COM, 1219 avenue Champs Blancs, 35510 Cesson-Sévigné — FRANCE

Where $\mathbf{x}_i$ is a signal from a training set, i.e. a block of pixels reshaped into a $N^2 \times 1$ vector, $\mathbf{X}_i = \mathbf{A} \cdot \mathbf{x}_i$ are the transformed coefficients using the transform $\mathbf{A}$ and $\mathbf{c}_i$ are the quantised transformed coefficients. $\mathbf{A}^T$ is the inverse transform, since $\mathbf{A}$ is chosen orthonormal.

The constraint in the cost function is the ℓ_0 norm of the coefficients, that is, the number of non-zero quantised transformed coefficients, also called the number of significant coefficients. The Lagrange multiplier λ of the constraint only depends on the quantisation accuracy applied to the coefficients, as proved in [7].

A solution to equation 1 can be found in two steps [7]. First, the optimal coefficients are obtained by hard-thresholding $\mathbf{X}_i$. Afterwards, the transform is updated given the hard-thresholded coefficients $\mathbf{c}_i$ and the $\mathbf{x}_i$. These two steps are repeated until convergence, with the design metric value being stable.

Although equation 1 outputs a non-separable transform, separable RDOTs can also be derived. In image coding, a two-dimensional separable transform is performed as:

$$\mathbf{X} = \mathbf{A}_v \cdot \left(\mathbf{A}_h \cdot \mathbf{x}^T\right)^T = \mathbf{A}_v \cdot \mathbf{x} \cdot \mathbf{A}_h^T \tag{2}$$

Consequently, equation 1 is updated as:

$$\mathbf{A}_{opt} = \arg\min_{\mathbf{A}} \sum_{\forall i} \min_{\mathbf{c}_i} \left(\|\mathbf{x}_i - \mathbf{A}_v^T \mathbf{c}_i \mathbf{A}_h\|_2^2 + \lambda \|\mathbf{c}_i\|_0 \right) \tag{3}$$

In order to learn a transform set specific to each IP mode, residuals coming from every prediction mode are used separately. These residuals come from a video encoding performed with the HEVC reference software (HM) using a set of sequences different from the HEVC test set.

For each prediction mode, an initial random classification of the residuals into $1+N$ groups is carried out. The first group is left to the HEVC core transforms and a RDOT is learnt on each remaining group. Based on the updated transforms, a reclassification of the residuals is performed using the design metric from equation 1. Those steps are iterated until the whole system becomes stable; the set of transforms is considered optimal then. See algorithm 1.

input : Residuals $\mathbf{x}$ from a given intra prediction mode
output: Set of N RDOTs A_n

Initial random classification into $1+N$ classes

while *!convergence* **do**
  for $n = 1$ **to** N **do**
    Learn a RDOT on Class$_n$ using equation 1
  end
  foreach *block* **x do**
    for $n = 0$ **to** N **do**
      $\delta_n = \|\mathbf{x} - \mathbf{A}_n^T \mathbf{c}\|^2 + \lambda \|\mathbf{c}\|_0$
    end
    $n^* = \arg\min_n(\delta_n)$
    Class$_{n^*}$.append (**x**)
  end
end

Algorithm 1: Multiple transform design

2.1. The use of separable and non-separable transforms

Non-separable transforms have frequently been discarded in favour of their separable counterparts in image and video coding. This is due to two reasons: the computational complexity and the storage requirements, which increase significantly when separable transforms are replaced with non-separable.

For non-separable transforms, the number of operations required for an $N \times N$ block is about N^4, whereas for a separable transform is reduced to $2N^3$. This amount of operations can further be reduced for transforms exhibiting symmetries such as the DCT-like transforms.

The storage requirements differ notably, as well. Assuming each transform coefficient is stored using 2 bytes, to store a non-separable transform, $2N^4$ bytes are needed. For two-dimensional separable transforms, $2 \cdot 2N^2 + N^2 = 5N^2$ are needed, instead. The second term in the separable transform equation represents the scanning pattern, which can be stored in one byte, such that the order of the coefficient is adapted to the context adaptive binary arithmetic coding (CABAC): the energy of the coded samples needs to be sorted in a descending order as in HEVC.

3. MODE-DEPENDENT TRANSFORM COMPETITION

The results from previous work in intra coding using mode-dependent transforms [5] and transform competition [6] independently encourage the combination of both, named mode-dependent transform competition (MDTC).

In order to validate the approach of using multiple transforms per IP mode, a first experiment has been carried out, where one additional transform is used to complement the HEVC core transforms. A flag has been used to indicate whether the additional transform has been chosen in the RDO loop.

The system extends HEVC for the 4×4 and 8×8 transform unit (TU) sizes. The encoder selects the best prediction mode / transform pair in the RDO loop: as such, there is a competition between the newly designed transforms and HEVC core transforms.

Results reported here use the coding configurations established by the Joint Collaborative Team on Video Coding (JCT-VC) standardisation group [9]. They consist in encoding the sequences at four quantisation parameter (QP) points (22, 27, 32, 37) and computing the average bit-rate reduction [10]. The experiments have been carried out in all intra (AI) and random access (RA) configurations, even though transform competition is only available for those blocks coded using intra predictions.

Regarding the complexity, the systems tested in this paper present significant improvements over the previous one from [6], which was 8 times more complex than the HEVC encoder. The current approach is only 2 times more complex.

Compression performance compared to HEVC is significantly better for both systems: they achieve a bit-rate reduction of 3.78% for the non-separable version, and 1.66% for the separable one.

Detailed results of this system are shown in table 1. As a reminder, non-separable systems from previous work achieved a bit-rate reduction of around 2% each one.

This simple experiment motivates the increase in the number of transforms in each IP mode to find out the amount of improvement that MDTC can provide. However, this increase of the number of transforms comes at a cost, notably in the storage and computational requirements: the results of this approach and the complexity requirements are discussed in the next section.

4. HIGH PERFORMANCE MDTC

This section presents the results of a system that combines mode-dependent transforms with an increased number of transforms for each prediction mode

Transform competition has been enabled for the 35 IP modes only for the 4×4 and 8×8 TU sizes. For this updated system, 16 transforms for 4×4 blocks and 32 for 8×8 blocks are designed.

Transform usage is indicated using a basic signalling scheme conforming the signalling used in section 3: a flag is used to inform whether the default HEVC transform has been used. In a negative case, the selected transform is signalled using a fixed length codeword, 4 bits for the 4×4 TUs sizes and 5 bits for 8×8 sizes.

The performances obtained with this higher complexity MDTC scheme have increased significantly: for the AI configuration, on average 7.1% compression gains are obtained with the non-separable version compared to HEVC, 4.1% are reported for the separable version. These average results are consistent over all the tested resolutions, although the improvement is lower for the higher resolution (class A), since TU sizes of 16×16 and 32×32 are more often used and competition has not been introduced for them.

The RA performance is lower than the one obtained in AI, as the proposed improvement only affects the intra modes: inter coding has not been modified with the changes proposed in this publication. Despite that, the improvement is significant, as improving the intra predicted blocks also provides a better reference for inter-predicted blocks.

In order to compare the visual impact of the proposed system, the best performing sequence has been used. The fact of having a bit-rate reduction of 25% makes it easier to spot differences and possible visual artefacts. The QP point used for comparison is 37, because both sequences present a bit-rate of around 3 Mbps, as displayed in figure 1. One can notice how the sparsity reduction affects the bit rate for the coding points: the transform signalling which requires a non negligible codeword length per TU is counterbalanced by the increase of the residual compactness.

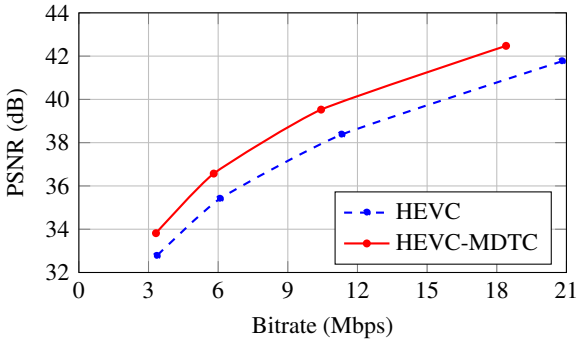
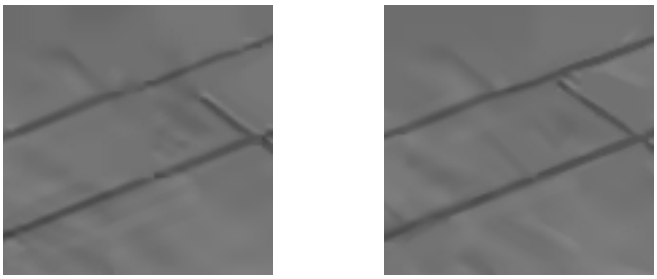


Fig. 1: BD-rate savings for BasketballDrill: -25.06%



(a) HEVC (b) HEVC-MDTC

Fig. 2: Visual comparison sample for BasketballDrill at QP37

As it can be observed in figure 2, this system, not only improves HEVC in textured and directional image regions, but also removes major coding artefacts.

A comparison between both systems in their separable and non-separable versions is presented in table 2. The number of transforms used affects, especially, the encoding time and the storage requirements, since more coding possibilities are explored. The augmented number of transforms highly increases the number of operations needed to find the best combination of block size, prediction mode and transform. Due to the increment of operations needed to find the best combination of block size, prediction mode and the augmented number of transforms, the separability has almost no impact on the encoding time. However, on the decoder side, using non-separable transforms comes at a high expense, around 30% can be observed, whereas complexity increases of about 5% when using separable transforms. Storage seems unreasonable when using a high number of transforms or non-separable transforms.

Figure 3 highlights the interest of increasing the number of additional transforms in each IP. The number of 8×8 transforms has been progressively increased and coding performance improve systematically. Better compression gains could be obtained by a further increase of the number of transforms. However, the complexity leads to unrealistic systems,

	Sequence	4×4: 1+1 — 8×8: 1+1				4×4: 1+16 — 8×8: 1+32			
		Y BD-rate (N-Sep)		Y BD-rate (Sep)		Y BD-rate (N-Sep)		Y BD-rate (Sep)	
		AI	RA	AI	RA	AI	RA	AI	RA
Class A (2560 × 1600)	NebutaFestival	-0.52%	-0.04%	-0.34%	-0.05%	-1.15%	-0.13%	-1.06%	-0.08%
	PeopleOnStreet	-2.76%	-1.16%	-1.40%	-0.48%	-5.65%	-2.27%	-4.21%	-1.49%
	SteamLocTrain	-0.46%	-0.36%	-0.36%	0.33%	-0.68%	0.03%	-0.60%	0.23%
	Traffic	-3.07%	-1.68%	-1.68%	-1.25%	-6.06%	-5.12%	-4.52%	-3.73%
Class B (1920 × 1080)	BasketballDrive	-3.15%	-1.76%	-1.17%	-0.35%	-5.52%	-2.36%	-3.22%	-0.72%
	BQTerrace	-5.11%	-2.83%	-1.80%	-1.08%	-9.22%	-4.93%	-4.70%	-2.65%
	Cactus	-3.76%	-2.42%	-1.95%	-1.15%	-10.92%	-7.68%	-5.22%	-3.08%
	Kimono1	-1.06%	-0.50%	-0.46%	-0.25%	-1.80%	-1.18%	-1.10%	-0.79%
	ParkScene	-2.20%	-1.50%	-1.68%	-1.09%	-5.27%	-3.69%	-4.61%	-3.22%
Class C (832 × 480)	BasketballDrill	-12.83%	-7.06%	-2.09%	-1.32%	-25.06%	-14.80%	-5.92%	-3.53%
	BQMall	-3.27%	-1.91%	-2.04%	-1.20%	-6.21%	-3.64%	-4.79%	-2.79%
	PartyScene	-3.31%	-2.29%	-2.18%	-1.40%	-6.19%	-4.30%	-4.88%	-3.21%
	RaceHorses	-3.99%	-2.16%	-1.71%	-0.70%	-6.75%	-3.24%	-4.25%	-1.59%
Class D (416 × 240)	BasketballPass	-3.71%	-1.97%	-1.68%	-0.82%	-6.15%	-3.14%	-3.91%	-1.82%
	BlowingBubbles	-4.21%	-2.35%	-1.96%	-1.21%	-6.73%	-3.95%	-4.30%	-2.69%
	BQSquare	-3.72%	-2.00%	-2.26%	-1.21%	-6.12%	-3.61%	-4.58%	-2.74%
	RaceHorses	-4.61%	-2.03%	-1.57%	-0.58%	-7.13%	-3.20%	-3.85%	-1.54%
Class E (1280 × 720)	FourPeople	-3.42%	-3.27%	-1.83%	-1.94%	-6.15%	-3.14%	-4.56%	-5.18%
	Johnny	-2.96%	-2.86%	-1.29%	-1.61%	-6.73%	-3.95%	-3.34%	-4.16%
	KristenAndSara	-3.49%	-3.26%	-1.34%	-1.75%	-6.12%	-3.61%	-3.82%	-4.47%
Class F (various resolutions)	BasketDrillText	-11.01%	-6.31%	-2.37%	-1.51%	-21.14%	-13.15%	-6.24%	-3.71%
	ChinaSpeed	-2.87%	-1.58%	-2.02%	-1.22%	-4.86%	-3.03%	-4.25%	-2.71%
	SlideEditing	-1.82%	-1.97%	-2.04%	-2.18%	-3.67%	-4.06%	-4.80%	-5.05%
	SlideShow	-3.49%	-2.96%	-2.53%	-2.32%	-6.03%	-5.91%	-5.66%	-5.61%
All sequences	**Overall**	**-3.78%**	**-2.36%**	**-1.66%**	**-1.10%**	**-7.10%**	**-4.67%**	**-4.10%**	**-2.76%**

Table 1: Compression gains for the different proposed MDTC configurations

	4×4: 1+1 8×8: 1+1		4×4: 1+16 8×8: 1+32	
	Sep	N-Sep	Sep	N-Sep
Y BD-rate	-1.66%	-3.78%	-4.10%	-7.10%
Enc. Time	×2	×2	×10	×10
Dec. Time	+3%	+30%	+5%	+30%
Storage	27 kB	298 kB	788 kB	9 MB

Table 2: Performance and complexity summary

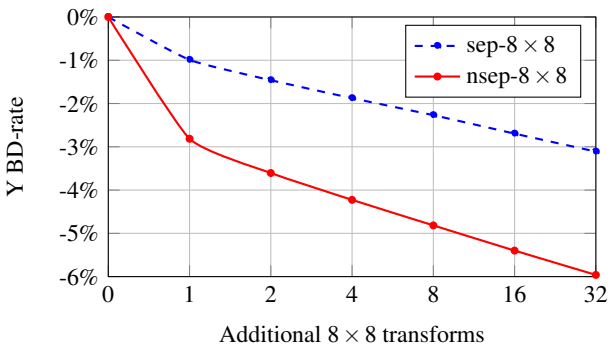


Fig. 3: Improvements over HEVC when the number of 8 × 8 transforms increases

especially for the encoding time and storage requirements.

5. CONCLUSIONS

This paper combines two approaches that provide extended competition for improved video coding: different transforms are designed per prediction mode to better adapt to the video signal statistics. The design of the transforms is based on a method based on previous work, and the level of performance obtained here demonstrates the validity of this method.

Two distinct systems were presented which report significant improvement over the state-of-the-art HEVC codec. Up to 7% of improvement is reported for the higher complexity configuration. Since the storage and algorithmic requirements for this system seem unrealistic, two alternative systems, using separable transforms, provide viable points in terms of ROM and complexity, they accommodate different trade-offs, for those systems 4% of compression improvement is provided for the higher complexity systems, 1.6% for the simpler coding scheme.

It has also been shown that different levels of performances can be achieved, depending on the complexity level. As a result, further work will address the complexity aspects: the storage requirements and the encoding complexity will be decreased as they represent the major drawbacks for the current designs.

6. REFERENCES

[1] G.J. Sullivan, J. Ohm, Woo-Jin Han, and T. Wiegand, "Overview of the High Efficiency Video Coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649–1668, 2012.

[2] Yan Ye and M. Karczewicz, "Improved H.264 intra coding based on bi-directional intra prediction, directional transform, and adaptive coefficient scanning," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, 2008, pp. 2116–2119.

[3] Chuohau Yeo, Yih Han Tan, Zhengguo Li, and S. Rahardja, "Mode-dependent fast separable klt for block-based intra coding," Tech. Rep. JCTVC-B024, ITU-T, Geneva, Switzerland, July 2010.

[4] A. Saxena and F.C. Fernandes, "Mode-dependent DCT/DST for intra prediction in video coding," Tech. Rep. JCTVC-D033, ITU-T, Guangzhou, China, October 2010.

[5] A. Arrufat, P. Philippe, and O. Déforges, "Non-separable mode dependent transforms for intra coding in HEVC," in *Visual Communications and Image Processing, 2014. IEEE Proceedings on*, 2014, pp. 61–64.

[6] A. Arrufat, P. Philippe, and O. Déforges, "Rate-distortion optimised transform competition for intra coding in HEVC," in *Visual Communications and Image Processing, 2014. IEEE Proceedings on*, 2014, pp. 73–76.

[7] O.G. Sezer, O. Harmanci, and O.G. Guleryuz, "Sparse orthonormal transforms for image compression," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, 2008, pp. 149–152.

[8] O. G. Sezer, *Data-driven transform optimization for next generation multimedia applications*, Ph.D. thesis, Georgia Institute of Technology, 2011.

[9] F. Bossen, "Common test conditions and software reference configurations," Tech. Rep. JCTVC-I1100, ITU-T, Geneva, Switzerland, May 2012.

[10] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," Tech. Rep. VCEG-M33, ITU-T, Austin, Texas, April 2001.